Unsupervised Nonlinear Spectral Unmixing Based on a Multilinear Mixing Model

Qi Wei, *Member, IEEE*, Marcus Chen, Jean-Yves Tourneret, *Senior Member, IEEE*, and Simon Godsill, *Member, IEEE*

***Abstract*— In the community of remote sensing, nonlinear mixture models have recently received particular attention in hyperspectral image processing. In this paper, we present a novel nonlinear spectral unmixing method following the recent multilinear mixing model of Heylen and Scheunders, which includes an infinite number of terms related to interactions between different endmembers. The proposed unmixing method is unsupervised in the sense that the endmembers are estimated jointly with the abundances and other parameters of interest, i.e., the transition probability of undergoing further interactions. Nonnegativity and sum-to-one constraints are imposed on abundances while only nonnegativity is considered for endmembers. The resulting unmixing problem is formulated as a constrained nonlinear optimization problem, which is solved by a block coordinate descent strategy, consisting of updating the endmembers, abundances, and transition probability iteratively. The proposed method is evaluated and compared with existing linear and nonlinear unmixing methods for both synthetic and real hyperspectral data sets acquired by the airborne visible/infrared imaging spectrometer sensor. The advantage of using nonlinear unmixing as opposed to linear unmixing is clearly shown in these examples.**

***Index Terms*— Block coordinate descent (BCD), gradient projection method, multilinear model, nonlinear unmixing (NLU).**

I. Introduction

SPECTRAL unmixing (SU) aims at decomposing a set of n multivariate measurements (or pixel vectors) $\mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^n]$ into a collection of m elementary signatures $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_m]$, usually referred to as *endmembers*, and estimating the relative proportions $\mathbf{A} = [\mathbf{a}^1, \ldots, \mathbf{a}^n]$ of these signatures, called *abundances*. SU has been advocated as a relevant multivariate analysis technique in various applicative areas, including remote sensing [2], planetology [3], microscopy [4], spectroscopy [5], and gene expression analysis [6]. In particular, a great interest has been demonstrated when analyzing multiband (e.g., hyperspectral) images, for instance for pixel classification [7], material quantification [8], and subpixel detection [9]. Due to the model inaccuracies, observation noise, environmental conditions, endmember variability, and data set size, SU is still a challenging ill-posed inverse problem. Plenty of mixing models and unmixing algorithms have been proposed to solve the SU problem and can be roughly classified into linear and nonlinear unmixing models [10]. In the following, we introduce the main linear mixture model (LMM) and nonlinear mixture model, the reasons for nonlinearity, and the existing solutions in this field.

Manuscript received April 14, 2016; revised December 18, 2016; accepted March 28, 2017. Date of publication May 25, 2017; date of current version July 20, 2017. This work was supported in part by DSO Singapore, in part by the HYPANEMA ANR Project under Grant ANR-12-BS03-003, and in part by the Thematic Trimester on Image Processing of the CIMI Labex, Toulouse, France, under Grant ANR-11-LABX-0040-CIMI and Program ANR-11-IDEX-0002-02. *(Corresponding author: Qi Wei.)*

Q. Wei is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: qi.wei@duke.edu).

M. Chen is with the DSO National Laboratories, Singapore 639798 (e-mail: marcuschen@pmail.ntu.edu.sg).

J.-Y. Tourneret is with the IRIT/INP-ENSEEIHT, University of Toulouse, 31071 Toulouse, France (e-mail: jean-yves.tourneret@enseeiht.fr).

S. Godsill is with the Department of Engineering, University of Cambridge, Cambridge CB2 1TN, U.K. (e-mail: sjg@eng.cam.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2693366

A. Linear Mixture Model

The LMM assumes that each image pixel is a linear combination of all the endmembers present in this pixel. The LMM is a simplified spectral mixture model that considers only first-order scattered photons by neglecting multiple photon interactions [11]. The LMM model has been widely used in the remote sensing community and can be expressed as [12]

$$\mathbf{x}^i = \mathbf{E}\mathbf{a}^i + \mathbf{n}^i \tag{1}$$

where

1) $\mathbf{x}^i$ is a $d \times 1$ vector representing the measured reflectance for the ith pixel;
2) $\mathbf{E} \in \mathbb{R}^{d \times m}$ is a nonnegative matrix whose columns $\mathbf{e}_1, \ldots, \mathbf{e}_m$ correspond to m endmember signatures and span the space where the data $\mathbf{x}^1, \ldots, \mathbf{x}^n$ reside;
3) $\mathbf{a}^i$ is a $m \times 1$ nonnegative vector, which includes the fractional abundances (coefficients) for the ith pixel (that sum to 1);
4) $\mathbf{n}^i \in \mathbb{R}^d$ is the additive Gaussian noise.

By arranging all pixels of the observed scenario lexicographically, the LMM model can be written as

$$\mathbf{X} = \mathbf{E}\mathbf{A} + \mathbf{N} \tag{2}$$

where $\mathbf{X} \in \mathbb{R}^{d \times n}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, and $\mathbf{N} \in \mathbb{R}^{d \times n}$ are the reflectance, abundance, and noise matrices, n is the number of observations, and d is the number of spectral bands.

The SU problem based on the LMM is generally formulated as the following constrained least squares problem:

$$\min_{\mathbf{a}^i} \|\mathbf{x}^i - \mathbf{E}\mathbf{a}^i\|_2^2 \text{ subject to (s.t.) } \mathbf{a}^i \geq 0 \text{ and } \mathbf{1}_m^T \mathbf{a}^i = 1 \tag{3}$$

where $\mathbf{1}_m$ is an $m \times 1$ vector with all ones. Using all observed data and matrix notations, the optimization problem can be written as

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{E}\mathbf{A}\|_F^2 \quad \text{s.t.} \quad \mathbf{A} \geq 0 \quad \text{and} \quad \mathbf{1}_m^T \mathbf{A} = \mathbf{1}_n^T \tag{4}$$

where $\mathbf{A} \geq 0$ has to be understood in the element-wise sense, meaning that all the coefficients of $\mathbf{A}$ are nonnegative. Note that $\|\cdot\|_F$ is the Frobenius norm, which is defined as

$$\|\mathbf{X}\|_F = \sqrt{\text{trace}(\mathbf{X}^H \mathbf{X})}$$

where $\mathbf{X}^H$ denotes the conjugate transpose of $\mathbf{X}$ and trace($\mathbf{M}$) is the trace of the matrix $\mathbf{M}$ [13].

B. Nonlinear Mixture Model

Due to its simple and intuitive physical interpretation as well as tractable estimation process, the LMM has been widely used for unmixing, and has shown interesting results in various applications. However, there exist many scenarios, involving intimate mixtures or multipath reflections, for which the LMM is not appropriate and can be advantageously replaced by a nonlinear mixture model [1], [14]–[16]. Nonlinear spectral mixing occurs in the presence of multiple reflections and transmissions from a surface. One notable example is the case of scenes with large geometrical structures such as buildings or trees, where shadowing and mutual illumination involve multiple light scattering effects. Another example is the case of mineral mixtures (also referred to as intimate mixtures), where an incoming light ray can interact many times with the different mineral grains, and the single interactions assumed in the LMM can even become relatively rare. Furthermore, the LMM only considers reflection and disregards optical transmission, which can become quite important in vegetation and mineral mixtures. To solve these problems, nonlinear mixture models have been proposed as interesting alternatives to overcome the inherent limitations of the LMM. These models include the Hapke model [17], the Fan *et al.* model [18], the generalized bilinear model (GBM) [19], [20], the linear-quadratic model [21], the polynomial postnonlinear mixture model [22], the multilinear mixing (MLM) model [1], and many others (see [23] for an intensive review).

In this paper, we focus on the recently proposed MLM model [1] mainly due to its generality to consider all orders of interactions between endmembers. The MLM model follows two basic assumptions.

1) A light ray incoming from the source will interact with at least one material.
2) In the ith pixel, after each interaction with a material, there is a finite probability of undergoing further interactions denoted by P_i, and thus the probability of escaping the scene and reaching the observer is $1 - P_i$.

More specifically, the nonlinear mixture model of [1] is formulated as

$$\begin{aligned}\mathbf{x}^i = (1 - P^i)\mathbf{y}^i + (1 - P^i)P^i(\mathbf{y}^i \odot \mathbf{y}^i) \\ + (1 - P^i)(P^i)^2(\mathbf{y}^i \odot \mathbf{y}^i \odot \mathbf{y}^i) \cdots\end{aligned} \tag{5}$$

where $\mathbf{x}^i$ represents the observed reflectance for the ith pixel, $\mathbf{y}^i = \mathbf{E}\mathbf{a}^i$ is the linear term used in the traditional LMM model, and $\odot$ represents the Hadamard entry-wise product. It is easy to find that all the orders of interactions between all the materials/endmembers have been considered and explicitly formulated in (5), which makes it flexible and generic for nonlinear mixed data.

To the best of our knowledge, the MLM model is the first nonlinear model that includes all orders of interactions by introducing only a single parameter P^i, which describes the probability of further interactions. Furthermore, the summation in (5) can be conveniently simplified as the following fixed-point equation:

$$\mathbf{x}^i = (1 - P^i)\mathbf{y}^i + P^i \mathbf{y}^i \odot \mathbf{x}^i. \tag{6}$$

Note that P^i is different from pixel to pixel. For more details concerning the derivation of the MLM, we refer the reader to [1]. To achieve nonlinear unmixing, Heylen and Scheunders [1] considered the following optimization problem:

$$\arg \min_{\{\mathbf{a}^i, P^i\}} \left\|\mathbf{x}^i - \frac{(1 - P^i)\mathbf{y}^i}{1 - P^i \mathbf{y}^i}\right\|_2^2 \quad \text{s.t.} \quad \mathbf{y}^i = \mathbf{E}\mathbf{a}^i. \tag{7}$$

The endmember matrix $\mathbf{E}$ was suggested to be estimated using vertex component analysis (VCA) in [24], which is one of the state-of-the-art endmember extraction methods, and to be fixed in the unmixing, leading to a supervised unmixing method. However, the VCA algorithm is based on the LMM model, which is different from the MLM. Furthermore, the optimization with respect to $\mathbf{a}^i$ and P^i is highly nonlinear and nonconvex, preventing a unique solution to be obtained.

To overcome the difficulties mentioned above, this paper considers two main modifications with respect to the method in [1]. First, the objective function is slightly changed from (7) in order to avoid its highly nonlinearity with respect to the parameters to be estimated. This modification significantly decreases the complexity of the optimization problem (7), which will be illustrated later. Second, instead of fixing the endmember matrix using VCA, the output of an endmember extraction algorithm (EEA) is used as an initialization of an algorithm, which estimates the endmember matrix jointly with the abundances and the transition probability, leading to an unsupervised nonlinear mixing strategy. Compared with [1], the proposed approach is blind and thus far more challenging.

II. Nonlinear Spectral Unmixing: A BCD Scheme

The nonlinear unmixing problem investigated in this paper can be formulated as the following optimization problem:

$$\arg \min_{\{\mathbf{a}^i, P^i\}_{i=1}^m, \mathbf{E}} L(\mathbf{E}, \mathbf{A}, \mathbf{P}) \tag{8}$$

$$\begin{aligned}\text{with} \quad L(\mathbf{E}, \mathbf{A}, \mathbf{P}) &= \sum_{i=1}^{n} \left\|\mathbf{x}^i - (1 - P^i)\mathbf{y}^i - P^i \mathbf{y}^i \odot \mathbf{x}^i\right\|_2^2 \\ \mathbf{y}^i &= \mathbf{E}\mathbf{a}^i \\ \mathbf{a}^i &\geq 0 \text{ and } \mathbf{1}_m^T \mathbf{a}^i = 1 \\ 0 &\leq \mathbf{E} \leq 1 \\ P^i &\leq 1.\end{aligned}$$

Compared with (7), the objective in (8) is easier to be optimized with respect to the abundances $\mathbf{A}$ and nonlinearity $\mathbf{P}$ due to the disappearance of fractional form. To solve the problem (8), we propose to update $\mathbf{a}^i$, $\mathbf{E}$, and $\mathbf{P}^i$ alternatively, using a block coordinate descent (BCD) strategy. Even though (8) is a nonconvex problem with respect to $\mathbf{a}^i$, $\mathbf{E}$, and $\mathbf{P}^i$ jointly, it is interesting to note that each subproblem turns out to be a convex problem that has a unique solution. The BCD algorithm is known to converge to a stationary point of the objective function to be optimized provided that this objective function has a unique minimum point with respect to each variable [25, Prop. 2.7.1], which is the case for the criterion in (8). Thus, the BCD algorithm introduced in this paper converges to a stationary point of (8). Note that the nonlinear unmixing problem (8) includes a linear unmixing (abundance estimation) step, an endmember extraction step, and a transition probability estimation step. To ease the notation, we omit the upper indices i for $\mathbf{a}$ and $\mathbf{P}$ hereafter as they can be updated pixel by pixel in parallel. It is worth noting that one popular strategy to overcome the nonconvexity is to use simulation-based methods such as Markov Chain Monte Carlo methods (see [26] for a recent review). Such an approach would be computationally intensive, but could potentially yield improvement in performance and better estimation of the uncertainty inherent in the problem. However, the major drawback of being computationally expensive for simulation-based methods prevents their effective use in this application.

A. Optimization With Respect to $\mathbf{a}$

The optimization with respect to $\mathbf{a}$ can be expressed as

$$\arg\min_{\mathbf{a}} \ \|\mathbf{x} - (1-P)\mathbf{E}\mathbf{a} - P\,(\mathbf{E}\mathbf{a}) \odot \mathbf{x}\|_2^2 \quad \text{s.t. } \mathbf{a} \geq 0 \ \text{ and } \ \mathbf{1}_m^T\mathbf{a} = 1. \tag{9}$$

Straightforward computations lead to the following equivalent optimization problem:

$$\arg\min_{\mathbf{a}} \ \|\mathbf{x} - \tilde{\mathbf{E}}\mathbf{a}\|_2^2 \ \text{ s.t. } \ \mathbf{a} \geq 0 \ \text{ and } \ \mathbf{1}_m^T\mathbf{a} = 1 \tag{10}$$

where $\tilde{\mathbf{E}} = \mathbf{E} \odot [(1-P)\mathbf{1}_{d\times m} + P\mathbf{x}\mathbf{1}_m^T]$. Thus, the optimization with respect to $\mathbf{a}$ becomes a standard fully constrained least squares problem with a modified endmember matrix $\tilde{\mathbf{E}}$. To solve this classical convex problem, there exist plenty of methods, e.g., active-set [27], alternating direction method of multipliers [28] (sometimes referred to as sparse unmixing by variable splitting and augmented Lagrangian [29]), and projection-based methods [30]. Instead of solving (10) exactly, we use the gradient projection method [31], [32] to decrease the objective function defined in (10). More specifically, by denoting $g(\mathbf{a}) = \|\mathbf{x} - \tilde{\mathbf{E}}\mathbf{a}\|_2^2$, (10) can be rewritten as

$$\arg\min_{\mathbf{a}} g(\mathbf{a}) \ \text{ s.t. } \ \mathbf{a} \in \mathcal{A} \tag{11}$$

where $\mathcal{A} = \{\mathbf{a} \in \mathbb{R}^m | \mathbf{a} \geq 0 \text{ and } \mathbf{1}_m^T\mathbf{a} = 1\}$. Thus, the gradient of the objective $g(\mathbf{a})$ with respect to $\mathbf{a}$ can be calculated as

$$\nabla_{\mathbf{a}} g(\mathbf{a}) = \tilde{\mathbf{E}}^T(\tilde{\mathbf{E}}\mathbf{a} - \mathbf{x}).$$

Note that the gradient projection method is different from the conventional gradient descent method in that each update after a move along the gradient direction $\nabla_{\mathbf{a}} g(\mathbf{a})$ is projected onto the convex set $\mathcal{A}$ to force all the updates to belong to the set of feasible solutions, that is

$$\mathbf{a} = \Pi_{\mathcal{A}}\left(\mathbf{a} - \gamma_{\mathbf{a}}\nabla_{\mathbf{a}} g(\mathbf{a})\right), \quad \varepsilon \leq \gamma_{\mathbf{a}} \leq 2/L_{\mathbf{a}} - \varepsilon \tag{12}$$

where $\Pi_{\mathcal{A}}$ denotes the projection operator onto $\mathcal{A}$, $\varepsilon \in]0, \min\{1, 1/L_{\mathbf{a}}\}[$ and $L_{\mathbf{a}} = \|\tilde{\mathbf{E}}^T\tilde{\mathbf{E}}\|_F$ is the Lipschitz constant of $\nabla_{\mathbf{a}} g(\mathbf{a})$. The projection onto the (canonical) simplex $\mathcal{A}$ can be achieved with a finite algorithm,[1] such as from Michelot [33], Duchi *et al.* [34], and Condat [35].

The motivation to use this gradient projection algorithm is twofold. First, the convergence of a gradient projection within a BCD scheme is guaranteed (see more details in [36]–[38]). Second, the update (12) is less computationally intensive than solving the optimization problem (10) exactly, which requires iterative updates. In this paper, the stepsize $\gamma_{\mathbf{a}}$ is fixed to $1/L_{\mathbf{a}}$ to ensure a sufficient decrease of the objective value per iteration. The updating scheme for $\mathbf{a}$ is summarized in Algorithm 1. The computational complexity to calculate the abundances for all pixels is of the order $\mathcal{O}(\max\{d, m\}nm)$.

Algorithm 1 Minimization With Respect to $\mathbf{a}$

```
/* Calculate the Modified Endmembers
   (Pixel-Wised)                          */
Ẽ ← E ⊙ ((1 − P)1_{d×m} + P x 1_m^T);
/* Calculate the Lipschitz constant       */
L_a ← ‖Ẽ^T Ẽ‖_F;
/* Gradient projection update             */
â ← Π_A (a − ∇_a g(a)/L_a);
Output: â
```

B. Optimization With Respect to P

The optimization with respect to P can be formulated as

$$\arg\min_{P} \|\mathbf{x} - (1-P)\mathbf{y} - P\mathbf{y} \odot \mathbf{x}\|_2^2 \quad \text{s.t. } \ P \leq 1. \tag{13}$$

Obviously, problem (13) is convex and admits the following closed-form solution:

$$\hat{P} = \Pi_{[-\infty,1]}\left(\frac{(\mathbf{y} - \mathbf{y}\odot\mathbf{x})^T(\mathbf{y} - \mathbf{x})}{\|\mathbf{y} - \mathbf{y}\odot\mathbf{x}\|_2^2}\right) \tag{14}$$

where $[-\infty, 1]$ corresponds to the constraint $P \leq 1$ and $\mathbf{y} = \mathbf{E}\mathbf{a}$. The computational complexity to calculate the probability P for all pixels is of the order $\mathcal{O}(nd)$.

C. Optimization With Respect to $\mathbf{E}$

The optimization of the objective function in (8) with respect to $\mathbf{E}$ can be formulated as

$$\arg\min_{\mathbf{E}} \|\mathbf{x} - (1-P)\mathbf{E}\mathbf{a} - P\,(\mathbf{E}\mathbf{a}) \odot \mathbf{x}\|_2^2 \quad \text{s.t. } 0 \leq \mathbf{E} \leq 1. \tag{15}$$

[1] A finite algorithm is an iterative algorithm that converges in a finite number of steps.

The above problem can be equivalently rewritten as

$$\arg\min_{\mathbf{E}} \left\|\mathbf{x} - \left(\mathbf{E} \odot \tilde{\mathbf{A}}\right)\mathbf{1}_m\right\|_2^2 \quad \text{s.t.} \quad 0 \leq \mathbf{E} \leq 1 \tag{16}$$

where $\tilde{\mathbf{A}} = ((1-P)\mathbf{1}_d + P\mathbf{x})\,\mathbf{a}^T$. Considering all the observed pixels leads to

$$\arg\min_{\mathbf{E}} f(\mathbf{E}) \quad \text{s.t.} \quad 0 \leq \mathbf{E} \leq 1 \tag{17}$$

where $f(\mathbf{E}) = \sum_{i=1}^{n} \|\mathbf{x}^i - (\mathbf{E} \odot \tilde{\mathbf{A}}^i)\mathbf{1}_m\|_2^2$ and $\tilde{\mathbf{A}}^i = ((1-P^i)\mathbf{1}_d + P^i\mathbf{x}^i)(\mathbf{a}^i)^T$. The gradient of the objective $f(\mathbf{E})$ can therefore be calculated as follows:

$$\begin{aligned}
\nabla_{\mathbf{E}} f(\mathbf{E}) &= \sum_{i=1}^{n} \left[\left(\mathbf{E} \odot \tilde{\mathbf{A}}^i\right)\mathbf{1}_m - \mathbf{x}^i\right]\mathbf{1}_m^T \odot \tilde{\mathbf{A}}^i \\
&= \sum_{i=1}^{n} \begin{bmatrix} \left(\mathbf{e}_1 \odot \tilde{\mathbf{a}}_1^i \odot \tilde{\mathbf{a}}_1^i + \cdots + \mathbf{e}_m \odot \tilde{\mathbf{a}}_m^i \odot \tilde{\mathbf{a}}_1^i\right)^T \\ \left(\mathbf{e}_1 \odot \tilde{\mathbf{a}}_1^i \odot \tilde{\mathbf{a}}_2^i + \cdots + \mathbf{e}_m \odot \tilde{\mathbf{a}}_m^i \odot \tilde{\mathbf{a}}_2^i\right)^T \\ \vdots \\ \left(\mathbf{e}_1 \odot \tilde{\mathbf{a}}_1^i \odot \tilde{\mathbf{a}}_m^i + \cdots + \mathbf{e}_m \odot \tilde{\mathbf{a}}_m^i \odot \tilde{\mathbf{a}}_m^i\right)^T \end{bmatrix}^T \\
&\quad - \sum_{i=1}^{n} \begin{bmatrix} \left(\mathbf{x}^i \odot \tilde{\mathbf{a}}_1^i\right)^T \\ \left(\mathbf{x}^i \odot \tilde{\mathbf{a}}_2^i\right)^T \\ \vdots \\ \left(\mathbf{x}^i \odot \tilde{\mathbf{a}}_m^i\right)^T \end{bmatrix}^T
\end{aligned} \tag{18}$$

where $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_m]$ and $\tilde{\mathbf{A}}^i = [\tilde{\mathbf{a}}_1^i, \ldots, \tilde{\mathbf{a}}_m^i]$. The second order derivative (Hessian matrix) of $f(\mathbf{E})$ with respect to $\mathbf{E}$ is a tensor and not easy to be expressed explicitly. Thanks to the Hadamard product, the second order derivative can be computed row by row. More specifically, for the ith row of $\mathbf{E}$, denoted by $\boldsymbol{\epsilon}^j (\in \mathbb{R}^{1\times m})$, we have

$$\nabla_{\boldsymbol{\epsilon}^j}^2 f(\boldsymbol{\epsilon}^j) = \sum_{i=1}^{n} \begin{bmatrix} \tilde{a}_{1,j}^i \tilde{a}_{1,j}^i & \cdots & \tilde{a}_{1,j}^i \tilde{a}_{m,j}^i \\ \tilde{a}_{2,j}^i \tilde{a}_{1,j}^i & \cdots & \tilde{a}_{2,j}^i \tilde{a}_{m,j}^i \\ \vdots & & \\ \tilde{a}_{m,j}^i \tilde{a}_{1,j}^i & \cdots & \tilde{a}_{m,j}^i \tilde{a}_{m,j}^i \end{bmatrix} \tag{19}$$

where $\tilde{a}_{l,j}^i$ represents the element of the matrix $\tilde{A}^i$ located in the lth row and in the jth column and $j = 1, \ldots, d$. Thus, the Hessian matrix (or Lipschitz constant) of $\nabla_{\mathbf{E}} f(\mathbf{E})$ can be computed row by row.

Similar to the update of $\mathbf{A}$, the gradient projection method can be implemented as follows:

$$\boldsymbol{\epsilon}^j = \Pi_{[0,1]^{1\times m}}\left\{\boldsymbol{\epsilon}^j - \gamma_{\boldsymbol{\epsilon}^j} \nabla_{\boldsymbol{\epsilon}^j} f(\boldsymbol{\epsilon}^j)\right\}, \quad \varepsilon \leq \gamma_{\boldsymbol{\epsilon}^j} \leq 2/L_{\boldsymbol{\epsilon}^j} - \varepsilon \tag{20}$$

where $\varepsilon \in]0, \min\{1, 1/L_{\boldsymbol{\epsilon}^j}\}[$ and $L_{\boldsymbol{\epsilon}^j} = \|\nabla_{\boldsymbol{\epsilon}^j}^2 f(\boldsymbol{\epsilon}^j)\|_F$. The computational complexity to calculate the endmember matrix $\mathbf{E}$ is of the order $\mathcal{O}(\max\{md, m^2\}n)$. In this paper, the stepsize $\gamma_{\boldsymbol{\epsilon}^j}$ is fixed to $1/L_{\boldsymbol{\epsilon}^j}$ to ensure a sufficient decrease of the objective value per iteration. The update of $\mathbf{E}$ is summarized in Algorithm 2.

Algorithm 2 Minimization With Respect to $\mathbf{E}$

Input: $\mathbf{E}$, $\mathbf{a}^{1:n}$, $\mathbf{x}^{1:n}$, $P^{1:n}$

/* Compute the temporary variable $\tilde{\mathbf{A}}^i$ */

1 $\tilde{\mathbf{A}}^i \leftarrow \left((1-P^i)\mathbf{1}_d + P^i\mathbf{x}^i\right)\mathbf{a}^{iT}$ for $i = 1, \ldots, n$;

/* Compute the gradient */

2 $\nabla_{\mathbf{E}} f(\mathbf{E}) \leftarrow$ Update $\nabla_{\mathbf{E}} f(\mathbf{E})$ cf. (18);

/* Compute the Lipschitz constants row by row */

3 $L_{\boldsymbol{\epsilon}^j} \leftarrow \left\|\nabla_{\boldsymbol{\epsilon}^j}^2 f(\boldsymbol{\epsilon}^j)\right\|_F$ for $j = 1, \ldots, d$;

/* Compute $\hat{\mathbf{E}}$ row by row */

4 $\hat{\mathbf{E}} \leftarrow$ Update each row of $\hat{\mathbf{E}}$ cf. (20);

Output: $\hat{\mathbf{E}}$

D. Summary

The proposed algorithm is summarized in Algorithm 3. Note that the updates of $\mathbf{a}_i$ and $\mathbf{P}_i$ can be implemented for all pixels in parallel, explaining why the updates of $\mathbf{a}_i$ and $\mathbf{P}_i$ are given in matrix form in lines 4 and 5 of Algorithm 3, where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ and $\mathbf{P} = [P_1, \ldots, P_n]$. Note that the joint estimation problem of $\mathbf{E}$, $\mathbf{A}$, and $\mathbf{P}$ is nonconvex and thus admits multiple local optima. Thus, in practice, any other prior information is encouraged to be integrated in the estimation problem to alleviate its ill-posedness. For example, if we simply fix the endmember matrix *a priori*, the optimization will consist of alternating between $\mathbf{A}$ and $\mathbf{P}$, leading to a supervised nonlinear unmixing method, similar to the method investigated in [1].

Algorithm 3 Unsupervised Nonlinear Unmixing Based on the Multilinear Mixture Model

Input: $\mathbf{X}$

/* Initialize $\mathbf{E}$, $\mathbf{P}$ */

1 $\mathbf{E}^{(0)} \leftarrow \text{VCA}(\mathbf{X})$;

2 $\mathbf{P}^{(0)} \leftarrow \mathbf{0}$;

3 **for** $t = 1, 2, \ldots$ **to** *stopping rule* **do**

/* Update $\mathbf{A}$ cf. Algo. 1 or (10) */

4 $\mathbf{A}^{(t)} \in \{\mathbf{A} |\, L(\mathbf{E}^{(t-1)}, \mathbf{A}, \mathbf{P}^{(t-1)}) \leq L(\mathbf{E}^{(t-1)}, \mathbf{A}^{(t-1)}, \mathbf{P}^{(t-1)})\}$;

/* Update $\mathbf{P}$ cf. (14) */

5 $\mathbf{P}^{(t)} \in \arg\min_{\mathbf{P}} L(\mathbf{E}^{(t-1)}, \mathbf{A}^{(t)}, \mathbf{P})$;

/* Update $\mathbf{E}$ cf. Algo. 2 */

6 $\mathbf{E}^{(t)} \in \{\mathbf{E} |\, L(\mathbf{E}, \mathbf{A}^{(t)}, \mathbf{P}^{(t)}) \leq L(\mathbf{E}^{(t-1)}, \mathbf{A}^{(t)}, \mathbf{P}^{(t)})\}$;

7 **end**

8 Set $\hat{\mathbf{A}} = \mathbf{A}^{(t)}$, $\hat{\mathbf{E}} = \mathbf{E}^{(t)}$ and $\hat{\mathbf{P}} = \mathbf{P}^{(t)}$;

Output: $\hat{\mathbf{A}}$, $\hat{\mathbf{E}}$ and $\hat{\mathbf{P}}$

E. Convergence Analysis

The convergence of the proposed nonlinear unmixing algorithm can be analyzed under the framework of the BCD method. More specifically, the proposed nonlinear unmixing algorithm contains gradient projection steps within a BCD strategy, whose convergence has been proved under convexity [36] and nonconvexity assumptions [37], [39] (see [38] for a recent review). Assuming that the objective function f is a

continuously differentiable convex function whose gradient is Lipschitz, the above method, referred to as block coordinate gradient projection method in [36], has been proved to have sub-linear rate of convergence. Bolte *et al.* [37] explored the convergence of the iterates in a more general framework, which is referred to as proximal alternating linearized minimization (PALM). They first gave a convergence proof for two blocks under nonconvex and nonsmooth assumptions and then generalized it for more than two blocks (see more details in [37, Th. 1 and Sec. 3.6]). When the objective function is nonconvex, the sequence of iterates generated by PALM is guaranteed to converge to a stationary point of the objective function instead of converging to its optimal value. In this nonlinear unmixing application, the optimization problem is obviously nonconvex due to the entanglement of $\mathbf{E}$ and $\mathbf{A}$, which can be regarded as an extended nonnegative matrix factorization. Thus, according to the above analysis, the sequence generated by Algorithm 3 converges to a stationary point of the objective function $L(\mathbf{E}, \mathbf{A}, \mathbf{P})$.

III. Experiments Using Synthetic and Real Data

This section studies the performance of the proposed unsupervised nonlinear unmixing method using both synthetic and real data. We compare this method with the existing state-of-the-art linear and nonlinear unmixing strategies, including LMM [29], Fan *et al.* [18], GBM [19], polynomial post-nonlinear model (PPNM) [22], and MLM [1]. All algorithms have been implemented using MATLAB R2016b on a MacPro computer with Intel Core i7 CPU at 2.70 GHz and 16GB RAM. Unmixing results have been evaluated using the figures of merit described in Section III-A. Several experiments have been conducted using synthetic data sets with controlled ground-truth. These experiments are studied in Section III-B. Two real data sets associated with different applications are then considered in Section III-C.

A. Performance Measures

To analyze the quality of the estimated results, we have considered the following normalized mean square errors (NMSEs):

$$\text{NMSE}_{\mathbf{A}} = -20 \log \frac{\|\widehat{\mathbf{A}} - \mathbf{A}\|_F}{\|\mathbf{A}\|_F}$$

$$\text{NMSE}_{\mathbf{E}} = -20 \log \frac{\|\widehat{\mathbf{E}} - \mathbf{E}\|_F}{\|\mathbf{E}\|_F}$$

$$\text{NMSE}_{\mathbf{P}} = -20 \log \frac{\|\widehat{\mathbf{P}} - \mathbf{P}\|_F}{\|\mathbf{P}\|_F}.$$

The larger (because of the negative symbol) the NMSEs, the better the quality of the estimation. Another quality index is the spectral angle mapper (SAM), which measures the spectral distortion between the actual and estimated endmembers. The SAM is defined as

$$\text{SAM}_{\mathbf{E}}(\mathbf{e}_n, \hat{\mathbf{e}}_n) = \arccos\left(\frac{\langle \mathbf{e}_n, \hat{\mathbf{e}}_n \rangle}{\|\mathbf{e}_n\|_2 \|\hat{\mathbf{e}}_n\|_2}\right).$$

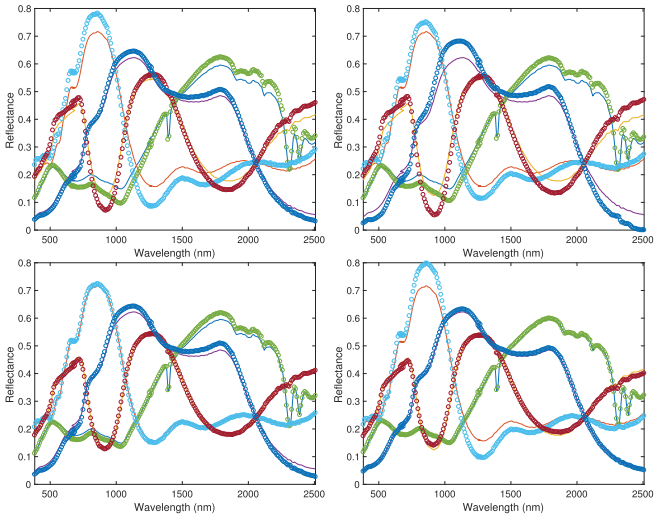

Fig. 1. Actual endmembers (solid lines) and their estimates (circles) with (Top left) VCA, (Top right) simplex identification via split augmented Lagrangian (SISAL), (Bottom left) N-FINDR, and (Bottom right) successive volume maximization (SVMAX).

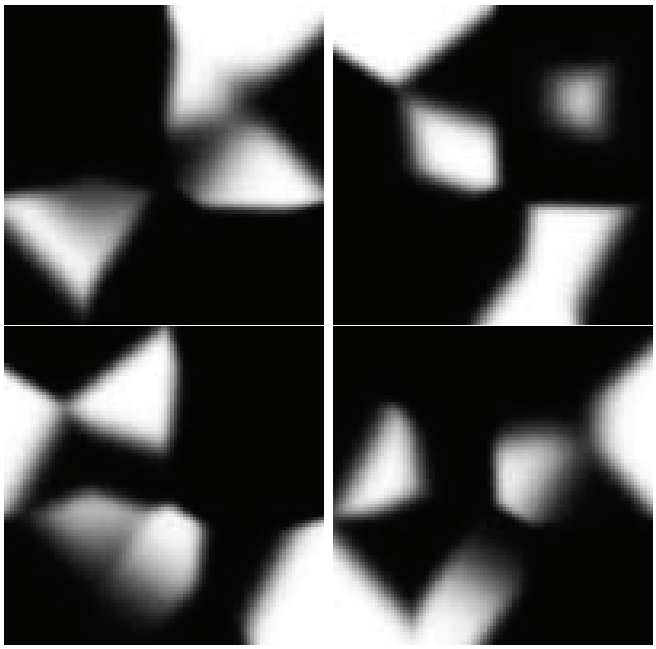

Fig. 2. Ground-truth for the four abundance maps.

The overall SAM is finally obtained by averaging the SAMs computed from all endmembers. Note that the value of SAM is expressed in degrees and thus belongs to $(-90, 90]$. The smaller the absolute value of SAM, the less important the spectral distortion.

To evaluate how well different models fit the observed data, the reconstruction error (RE) is computed as

$$\text{RE} = \|\hat{\mathbf{X}} - \mathbf{X}\|_F \tag{21}$$

where $\hat{\mathbf{X}}$ is the reconstructed data from the observation model. The smaller RE, the better the model fits the data.

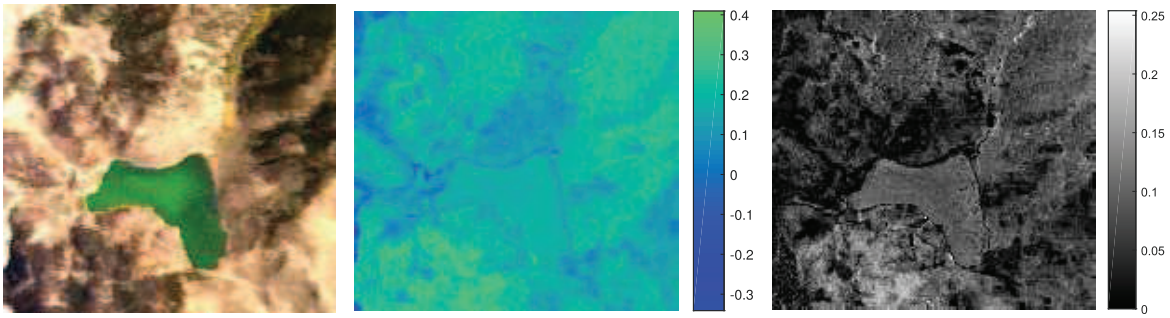

Fig. 3. (Left) Tahoe data set. (Middle) Estimated map $\hat{\mathbf{P}}$. (Right) Sum of absolute differences between abundance maps estimated by LU and NLU.

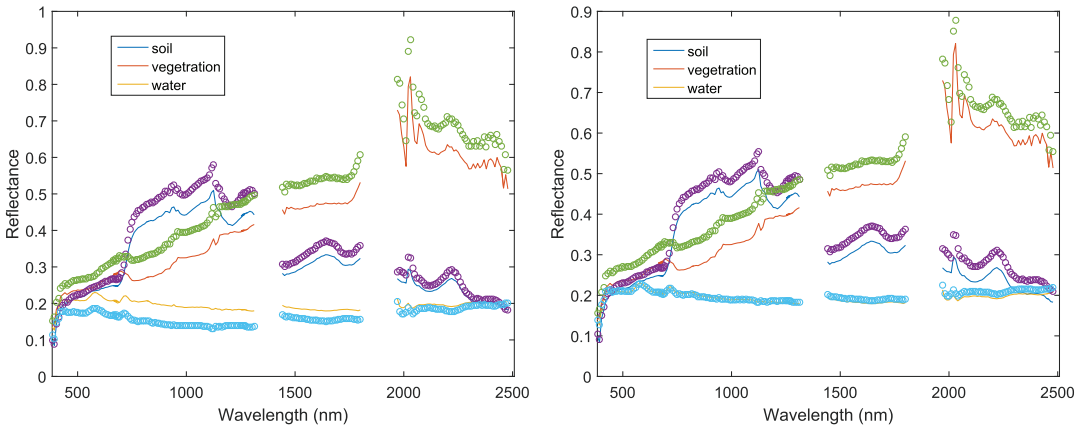

Fig. 4. Tahoe data set: estimated endmember signatures of soil, vegetation, and water using (Left) VCA (solid) and LU (circle) and (Right) VCA (solid) and NLU (circle).

B. Synthetic Data

In order to build the endmember matrix $\mathbf{E}$, we have randomly selected four spectral signatures from the United States Geological Survey Digital Spectral Library.[2] In this experiment, the number of endmembers is fixed to $m = 4$ and the reflectance spectra have $L = 224$ spectral bands ranging from 383 to 2508nm. The ground-truth of the endmembers is displayed with solid lines in Fig. 1. The abundance matrix $\mathbf{A}$ has been generated by drawing vectors distributed according to a Dirichlet distribution in the simplex $\mathcal{A}$ defined by the nonnegativity and sum-to-one constraints as in [40]. The generated ground-truth of the abundance maps is displayed in Fig. 2.

1) Data Models: To generate the synthetic data, several linear and unmixing models, e.g., LMM [29], Fan *et al.* [18], GBM [19], PPNM [22], MLM [1], and the proposed MLM (denoted by MLMp in the following section) have been used. Note that the MLM and MLMp models are slightly different following (7) and (8), respectively. For the GBM, the nonlinearity parameter γ was drawn randomly following a uniform distribution in the interval $[0,\ 1]$. For the PPNM, the nonlinearity parameter b was set to 0.25. For MLM and MLMp, we followed the strategy advocated in [1], i.e., the values of P for all pixels were generated by drawing samples randomly from a half-normal distribution (a truncated normal distribution on $\mathbb{R}^+$) with $\sigma = 0.3$. Values of P larger than one were set to zero, which led to mostly small values for $\mathbf{P}$ exemplifying a more realistic scenario. The performance of these algorithms was then evaluated on a synthetic data set of size $100 \times 100 \times 224$ with a signal-to-noise ratio (SNR) = 40dB.

[2]http://speclab.cr.usgs.gov/spectral.lib06/

TABLE I

UNMIXING PERFORMANCE OF PROPOSED ALGORITHM WITH DIFFERENT INITIALIZATIONS: $\text{NMSE}_{\mathbf{A}}$ (IN dB), $\text{NMSE}_{\mathbf{E}}$ (IN dB), $\text{SAM}_{\mathbf{E}}$ (IN DEGREES), RE, AND TIME (IN SECONDS)

Initial	$\text{NMSE}_{\mathbf{A}}$	$\text{NMSE}_{\mathbf{E}}$	$\text{SAM}_{\mathbf{E}}$	$\text{NMSE}_{\mathbf{P}}$	RE	time
VCA	48.58	49.99	0.047	33.39	5.24	149.3
SISAL	16.40	21.52	4.31	32.94	5.23	218.9
N-FINDR	23.27	30.19	1.19	33.87	5.25	130.7
SVMAX	22.24	25.95	1.78	27.09	5.36	187.9

2) Stopping Rule: As all the constraints associated with the endmembers and abundances are guaranteed to be satisfied at each update, the main issue after several updates is to analyze the value of the objective function. The stopping rule used in our experiments is defined as

$$\frac{L(\mathbf{E}^{(t)}, \mathbf{A}^{(t)}, \mathbf{P}^{(t)}) - L(\mathbf{E}^{(t-1)}, \mathbf{A}^{(t-1)}, \mathbf{P}^{(t-1)})}{L(\mathbf{E}^{(t-1)}, \mathbf{A}^{(t-1)}, \mathbf{P}^{(t-1)})} < \eta$$

where η was fixed to 10^{-4} by cross validation.

3) Initialization: For the proposed method, the initializations of $\mathbf{P}$, $\mathbf{A}$, and $\mathbf{E}$ are necessary. Following the strategy in [1], the nonlinearity matrix $\mathbf{P}$ was initialized with a zero matrix and $\mathbf{A}$ with the LMM unmixing result. Due to the highly nonconvexity of the nonlinear unmixing problem, the initialization of the endmembers is critical. To test the dispersion of the converged stationary points, the endmember matrix $\mathbf{E}$ was initialized using the outputs of the different

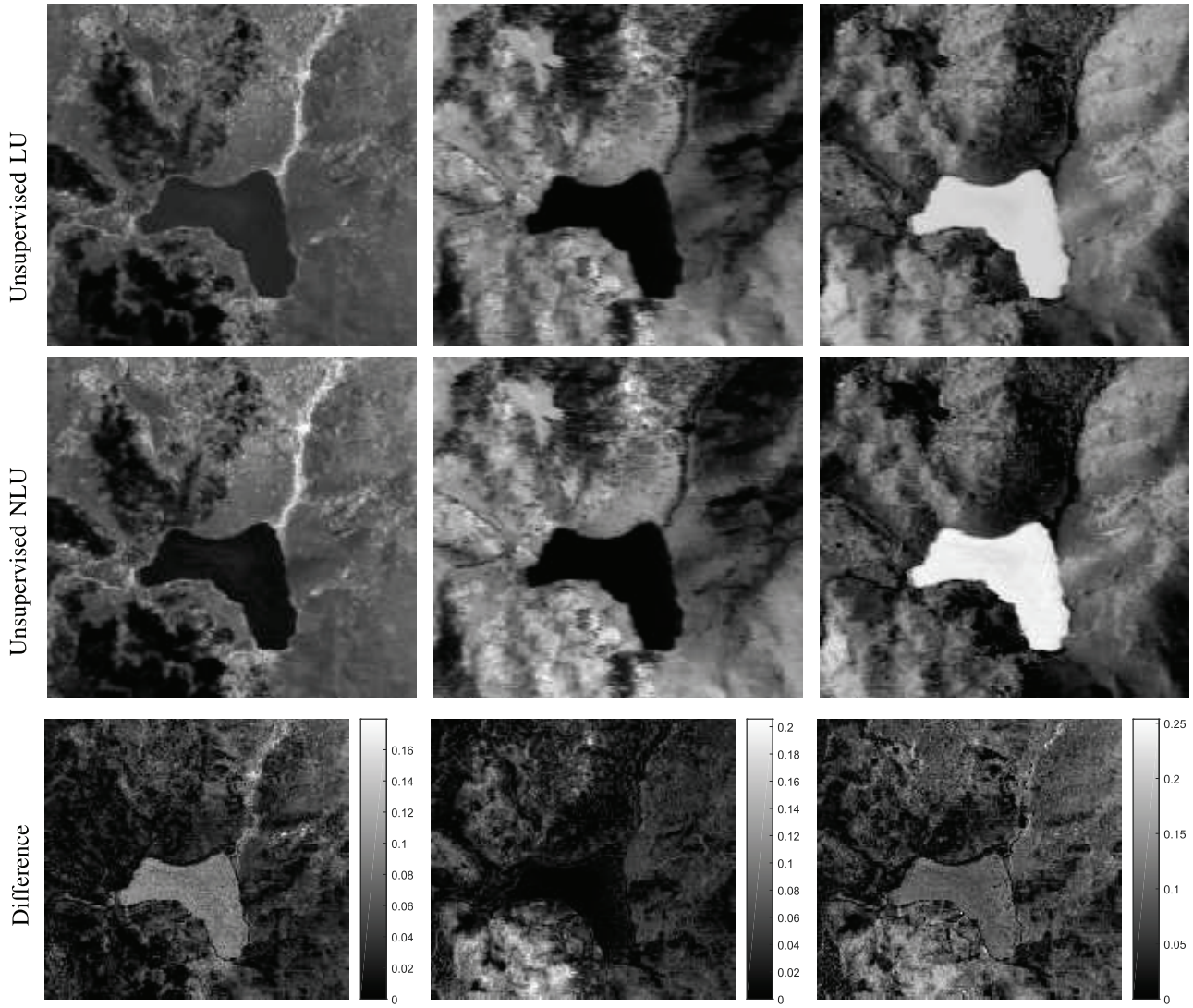

Fig. 5. Estimated abundance maps with LU and NLU methods. (Left to right) Soil, vegetation, and water. The abundances are linearly stretched between 0 (black) and 1 (white).

EEAs, e.g., resulting from VCA [24], SISAL [41], N-FINDR [42], and SVMAX [43]. The mixing model used in these experiments is the MLMp model. The reference endmembers and their estimates provided by the different EEAs are displayed in Fig. 1. Moreover, more quantitative unmixing results are reported in Table I. The estimates of $\mathbf{E}$, $\mathbf{A}$, and $\mathbf{P}$ clearly converge to different points while the REs are quite similar due to the nonconvexity of the unmixing problem. Among these initializations, VCA, one of the state-of-the-art endmember extraction methods, provides the best performance with all metrics with a reasonable computation time, demonstrating its superiority in extracting endmembers for this application.

4) Compare Different Models: To evaluate how the different models can fit the observed data, we generated synthetic data sets and unmixed them with different models. More specifically, we want to evaluate the unmixing performance when the unmixing model is different from the actual mixing model. Note that the endmembers were fixed to their ground-truth values for these experiments. The abundance maps and nonlinear parameters were estimated by minimizing the objective function under different model assumptions. Using these estimated abundances and parameters, we reconstructed all the pixels and the unmixing performance was evaluated using the

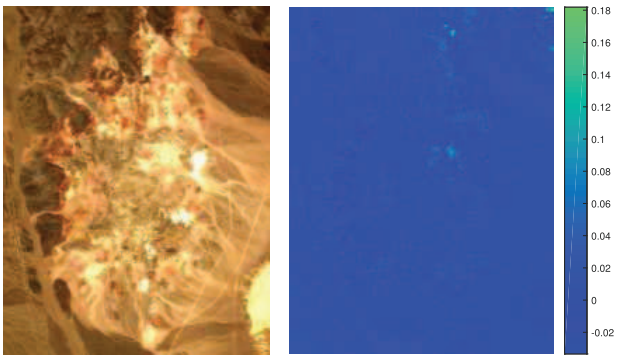

Fig. 6. (Left) Cuprite data set. (Right) Estimated map $\hat{\mathbf{P}}$.

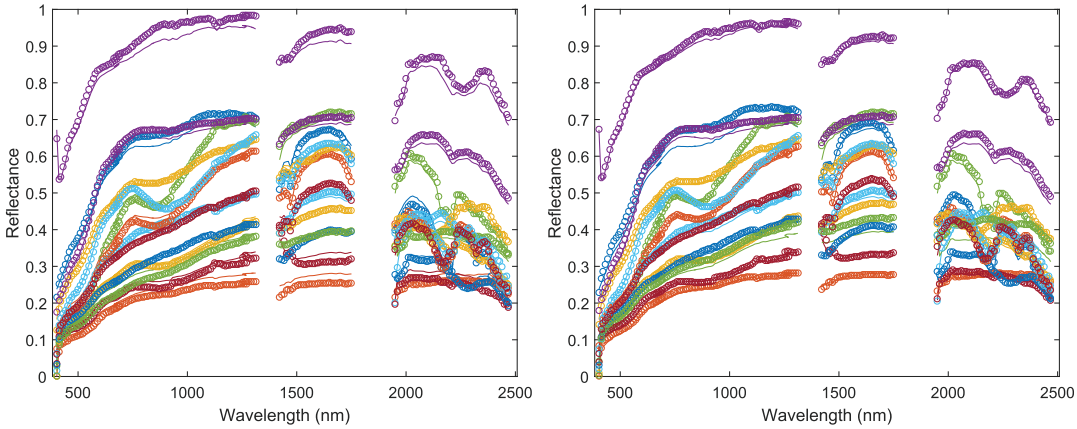

Fig. 7. Cuprite data set: extracted endmember signatures using (Left) VCA (solid) and LU (circle) and (Right) VCA (solid) and NLU (circle).

TABLE II
RE OF NLU WITH DIFFERENT MODELS

		unmixing model					
		LMM	Fan	GBM	PPNM	MLM	MLMp
mixing model	LMM	5.42	26.49	5.41	5.41	5.41	5.41
	Fan	28.02	5.61	5.61	6.14	11.74	11.74
	GBM	16.26	15.38	5.50	5.71	8.02	8.02
	PPNM	60.49	50.30	51.64	6.02	28.76	28.78
	MLM	110.54	128.82	110.54	41.42	4.61	4.61
	MLMp	110.56	128.84	110.56	41.48	5.07	5.07
time		16.85	17.27	30.31	17.16	17.50	11.61

TABLE III
$\text{NMSE}_{\mathbf{A}}$ OF NLU WITH DIFFERENT MODELS

		unmixing model					
		LMM	Fan	GBM	PPNM	MLM	MLMp
mixing model	LMM	52.55	25.74	51.98	51.63	52.42	52.53
	Fan	28.32	53.33	52.77	32.70	32.32	32.18
	GBM	32.76	30.92	47.78	37.01	36.29	36.15
	PPNM	23.02	27.44	27.66	51.94	25.79	25.57
	MLM	15.32	15.00	15.32	12.36	42.04	39.07
	MLMp	15.32	15.00	15.32	12.36	40.92	42.54
time		16.85	17.27	30.31	17.16	17.50	11.61

values of RE and $\text{NMSE}_{\mathbf{A}}$. The results are summarized in Tables II and III, where each row corresponds to a given data set and each column is associated with an unmixing model. The small diagonal elements of Table II are due to the fact that every unmixing model explains the data generated from the same model quite well. The first row of Table II also shows that most nonlinear unmixing methods perform well for linearly mixed data. The first column of Table II reporting the performance of linear unmixing shows that LMM fails to explain nonlinearly mixed data in most cases, as expected. The last two columns of Table II show that MLM and MLMp fit quite well (and similarly) all the mixed data (slightly worse for PPNM data but already much better than other linear or nonlinear models). Table III shows the values of the $\text{RMSE}_{\mathbf{A}}$ related to the abundances of the generative model. As the nonlinearity structure differs in the mixing and unmixing models, the values of $\text{RMSE}_{\mathbf{A}}$ indicate how generic the unmixing model is. The relatively large values obtained in the last two columns of Table III show the advantage of the MLM model when compared with other nonlinear models.

C. Real Data Sets

1) Tahoe Data Set: In this experiment, we consider an hyperspectral (HS) image of size $161 \times 174 \times 224$ acquired over Tahoe, located along the border between California and Nevada, on October 13, 2015, by the JPL/NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS).[3] This image was initially composed of 224 bands that have been reduced to 192 bands ($d = 192$) after removing the water vapor absorption bands as well as the highly noisy bands. The spatial resolution of this HS image is around 20 m per pixel and it mainly contains water, soil, and vegetation. A composite color image of the scene of interest is shown in the left of Fig. 3. In this experiment, the number of endmembers was fixed to be $m = 3$ according to our available prior knowledge. The three endmembers were estimated using VCA and are shown with solid lines in Fig. 4. As in Section III-B, the matrix $\mathbf{P}$ was initialized to $\mathbf{P} = \mathbf{0}$ and the abundances were initialized with the results of the LMM.

Note that fixing $\mathbf{P} = \mathbf{0}$ leads to an unsupervised LMM SU method. By fixing the endmember matrix $\mathbf{E}$ to the VCA estimates, we obtain the supervised versions of the algorithms for linear and nonlinear unmixing. To simplify the notations, linear and nonlinear unmixing are referred to as LU and NLU thereafter. Note that the supervised NLU can be regarded as a variant of Heylen's method in [1] with slightly different objectives. The NLU and LU (obtained by fixing $\mathbf{P} = \mathbf{0}$) were implemented to process the observed image. As there is no ground-truth for this image, unmixing results are first studied qualitatively by displaying the endmembers and abundances. Fig. 4 shows that the estimated signatures of soil and vegetation using LU and NLU are similar, while those of water are slightly different objective functions. The corresponding abundance maps obtained by LU and NLU shown in the first two rows of Fig. 5 (the color scales are exactly the same) are globally similar. However, some differences can be

[3] http://aviris.jpl.nasa.gov/

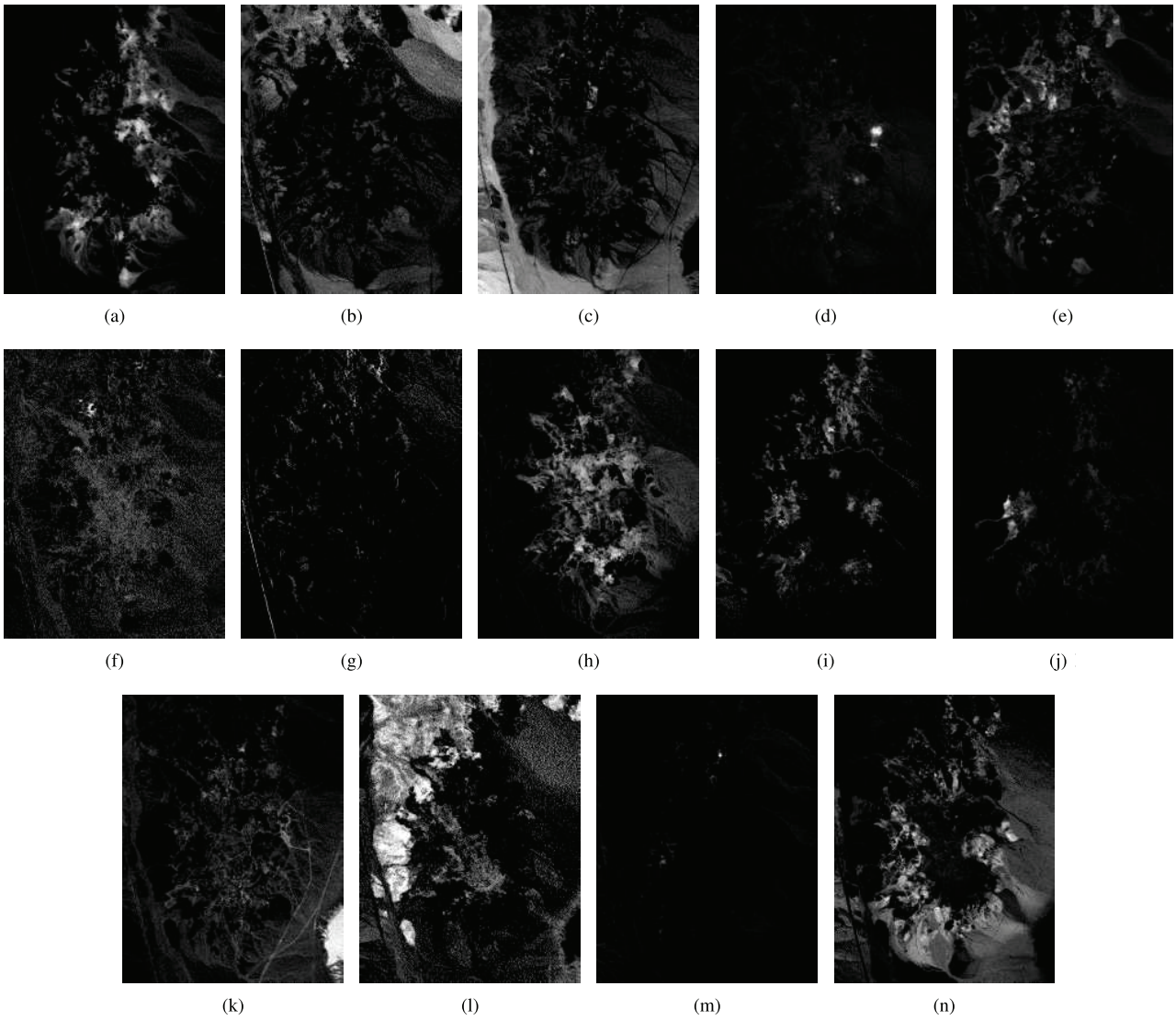

Fig. 8. Cuprite data set: abundance maps estimated by NLU. All abundances are linearly stretched between 0 (black) and 1 (white). (a) Alunite. (b) Sphene. (c) Andradite. (d) Muscovite. (e) Kaolinite #1. (f) Nontronite. (g) Kaolinite #4. (h) Kaolinite #3. (i) Pyrope #2. (j) Buddingtonite. (k) Montmorillonite. (l) Pyrope #1. (m) Kaolinite #2. (n) Dumortierite.

observed in the last row of Fig. 5. It is interesting to note that the abundance maps of NLU have larger contrast than the ones obtained with LU, which is usually expected for good unmixing results. For example, the lake part is expected to have small values in the abundance map of soil and big values in the abundance map of water. As shown in the first and second rows of Fig. 5, the lake part in NLU is darker than the one in LU for the soil map and brighter than the one in LU for the water map, showing that the results obtained by NLU are more reasonable. In order to appreciate the interest of using a nonlinear unmixing strategy, Fig. 3 shows the estimated map of $\mathbf{P}$ (in the middle), which reflects the distribution of nonlinearity and the sum of absolute differences between abundance maps estimated by LU and NLU. It is interesting to note that the areas with large differences in abundance maps are in good agreement with the areas associated with large values of $\mathbf{P}$. In these areas, nonlinear effects can be expected as there exist multiple interactions between endmembers in the ridge of mountain, in the shadow of mountain, in the lake, etc.

TABLE IV
RE AND TIME (IN SECONDS) OF DIFFERENT MODELS FOR TAHOE DATA SET

model	LMM	Fan	GBM	PPNM	MLM	MLMp
RE	29.38	30.90	28.61	26.65	26.01	25.53
time	40.32	40.39	64.35	59.26	45.26	330.31

As in Section III-B, other linear and nonlinear unmixing methods have also been used to unmix this data set. The corresponding RE and running time are reported in Table IV. Note that the main difference between MLM and MLMp here is that the endmember signatures are fixed for MLM but updated for MLMp. This explains why smaller RE is obtained for MLMp at the price of a higher computation cost.

2) Cuprite Data Set: This section investigates the performance of the proposed NLU method for unmixing the well-

TABLE V

RE AND TIME (IN S) OF DIFFERENT MODELS FOR CUPRITE DATA SET

model	LMM	Fan	GBM	PPNM	MLM	MLMp
RE	24.48	26.07	\	22.92	19.80	18.38
time	247.6	330.1	\	260.4	270.2	650.0

known Cuprite HS image. This image, which has received a lot of interest in the remote sensing and geoscience literature, was acquired over Cuprite field by AVIRIS. It corresponds to a mining area in southern Nevada composed of several minerals and some vegetation, located approximately 200 km northwest of Las Vegas. The image considered in this experiment consists of 250×191 pixels with $n_\lambda = 188$ spectral bands obtained after removing the water vapor absorption bands. A composite color image of the scene of interest is shown in the left of Fig. 6. According to [24], the number of endmembers was set to $m = 14$. The estimated endmember signatures from VCA, LU, and NLU are displayed in Fig. 7 and the $m = 14$ corresponding abundance maps estimated using NLU are shown in Fig. 8. The estimated map for $\mathbf{P}$ obtained with NLU is shown in the right of Fig. 6. Nonlinear effects can be expected in this scenario as there exist intimate mixtures, in which a light ray can interact many times with the different mineral grains. Comparisons with other linear and nonlinear unmixing methods are summarized in Table V. Note that the GBM model fails, i.e., does not converge in a reasonable time, to unmix this data set because of the overfitting problem caused by the relatively large number of endmembers. For the other models, the MLMp method provides the smallest RE (at the price of a higher computational cost), which is consistent with the results in Section III-B.

IV. CONCLUSION AND PERSPECTIVES

This paper studied a new unsupervised nonlinear SU method based on a recent multilinear model. The nonlinear unmixing problem was formulated as a constrained optimization problem with respect to the endmembers, abundances, and pixel-dependant transition probabilities. A gradient projection within BCD method was then proposed to estimate these three sets of variables jointly. Each step of the proposed method was carefully addressed to guarantee the convergence of the proposed algorithm to a stationary point of the objective function. Experiments implemented on both synthetic and real data sets confirmed that the multilinear model allows us to detect and analyze the nonlinearities present in a hyperspectral image. Comparisons with other linear and nonlinear unmixing methods also demonstrated the superiority of proposed algorithm. Further work will be devoted to investigate a similar nonlinear unmixing algorithm accounting for spatial correlations for the abundances and the transition probabilities as in [44] and [45]. Validating the nonlinear unmixing results in other practical applications, e.g., involving intimate mixtures or reflectances acquired after multipath, is also of interest.

ACKNOWLEDGMENTS

The authors would like to thank R. Heylen for insightful discussions on the design of the multilinear model.

REFERENCES

[1] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 240–251, Jan. 2016.

[2] A. Averbuch, M. Zheludev, and V. Zheludev, "Unmixing and target recognition in airborne hyper-spectral images," *Earth Sci. Res.*, vol. 1, no. 2, pp. 200–228, 2012.

[3] K. E. Themelis, F. Schmidt, O. Sykioti, A. A. Rontogiannis, K. D. Koutroumbas, and I. A. Daglis, "On the unmixing of MEx/OMEGA hyperspectral data," *Planetary Space Sci.*, vol. 68, no. 1, pp. 34–41, 2012.

[4] N. Dobigeon and N. Brun, "Spectral mixture analysis of EELS spectrum-images," *Ultramicroscopy*, vol. 120, pp. 25–34, Sep. 2012.

[5] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Process.*, vol. 89, no. 12, pp. 2657–2669, 2009.

[6] C. Bazot, N. Dobigeon, J.-Y. Tourneret, A. K. Zaas, G. S. Ginsburg, and A. O. Hero, III, "Unsupervised Bayesian linear unmixing of gene expression microarrays," *BMC Bioinform.*, vol. 14, no. 1, p. 99, Mar. 2013.

[7] C.-I. Chang, X.-L. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.

[8] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2601–2616, Sep. 2006.

[9] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.

[10] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[11] N. Yokoya, J. Chanussot, and A. Iwasaki, "Hyperspectral and multispectral data fusion based on nonlinear unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Shanghai, China, Jun. 2012, pp. 1–4.

[12] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[14] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2442–2453, May 2013.

[15] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.

[16] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

[17] B. Hapke, "Bidirectional reflectance spectroscopy: 1. Theory," *J. Geophys. Res.*, vol. 86, no. B4, pp. 3039–3054, Apr. 1981.

[18] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Int. J. Remote Sens.*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.

[19] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[20] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 2, pp. 1430–1437, Feb. 2014.

[21] I. Meganem, P. Déliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear–quadratic mixing model for reflectances in urban environments," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 544–558, Jan. 2014.

[22] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 21, no. 6, pp. 3017–3025, Jun. 2012.

[23] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.
[24] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[25] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[26] M. Pereyra *et al.*, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 224–241, Mar. 2016.
[27] D. C. Heinz and C.-I. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.
[28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[29] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process., Evol. Remote Sens. (WHISPERS)*, Reykjavik, Iceland, Jun. 2010, pp. 1–4.
[30] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret. (2015). "Fast spectral unmixing based on Dykstra's alternating projection." Unpublished paper, 2015. [Online]. Available: http://arxiv.org/abs/1505.01740
[31] P. H. Calamai and J. J. Moré, "Projected gradient methods for linearly constrained problems," *Math. Program.*, vol. 39, no. 1, pp. 93–116, 1987.
[32] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.
[33] C. Michelot, "A finite algorithm for finding the projection of a point onto the canonical simplex of $\mathbb{R}^n$," *J. Optim. Theory Appl.*, vol. 50, no. 1, pp. 195–200, 1986.
[34] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the ℓ_1 for learning in high dimensions," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Helsinki, Finland, 2008, pp. 272–279.
[35] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Math. Program.*, vol. 158, no. 1, pp. 575–585, 2016, doi: 10.1007/s10107-015-0946-6.
[36] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2037–2060, 2013.
[37] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, nos. 1–2, pp. 459–494, 2014.
[38] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, 2015.
[39] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward–backward algorithm," *J. Global Optim.*, vol. 66, no. 3, pp. 457–485, 2016. [Online]. Available: http://dx.doi.10.1007/s10898-016-0405-9
[40] E. Chouzenoux, M. Legendre, S. Moussaoui, and J. Idier, "Fast constrained least squares spectral unmixing using primal-dual interior-point optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 59–69, Jan. 2014.
[41] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Grenoble, France, Aug. 2009, pp. 1–4.
[42] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE* 3753, Imaging Spectrometry V, 266, Oct. 1999.
[43] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, vol. 3. Boston, MA, USA, Jul. 2008, pp. III-250–III-253.
[44] O. Eches, J. A. Benediktsson, N. Dobigeon, and J.-Y. Tourneret, "Adaptive Markov random fields for joint unmixing and segmentation of hyperspectral images," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 5–16, Jan. 2013.
[45] P. Chen, J. D. B. Nelson, and J.-Y. Tourneret, "Toward a sparse Bayesian Markov random field approach to hyperspectral unmixing and classification," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 426–438, Jan. 2017.

Qi Wei (S'13–M'15) was born in Shanxi, China, in 1989. He received the bachelor's degree in electrical engineering from Beihang University, Beijing, China, in 2010, and the Ph.D. degree in signal and image processing from the École Nationale Supérieure d'Électrotechnique, d'Électronique, d'Informatique, d'Hydraulique et des Télécommunications, National Polytechnic Institute of Toulouse, University of Toulouse, Toulouse, France, in 2015.

From 2015 to 2016, he was a Research Associate with the Signal Processing Laboratory, Department of Engineering, University of Cambridge, Cambidge, U.K., where he was involved in multiband image processing. He has been a Research Associate under the guidance of Prof. L. Carin with Duke University, Durham, NC, USA, since 2016. His research interests include machine/deep learning, computer vision/image processing, and Bayesian statistical inference.

Dr. Wei's doctoral thesis *Bayesian Fusion of Multi-band Images: A Powerful Tool for Super-Resolution* was rated as one of the best theses (awarded Prix Léopold Escande) with the University of Toulouse, in 2015.

Marcus Chen received the B.S. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, the M.S. degree in electrical engineering from Stanford University, Stanford, CA, USA, and the Ph.D. degree from the School of Computer Engineering, Nanyang Technological University, Singapore.

He is currently a Senior Member of the Technical Staff with DSO National Laboratories, Singapore. His research interests include remote sensing, the efficient learning of high dimensional data, and pattern recognition.

Jean-Yves Tourneret (SM'08) received the Ingénieur degree in electrical engineering from the École Nationale Supérieure d'Électrotechnique, d'Électronique, d'Informatique, d'Hydraulique et des (ENSEEIHT), Toulouse, France, in 1989, and the Ph.D. degree from the National Polytechnic Institute of Toulouse-ENSEEIHT, in 1992. He is currently a Professor with the ENSEEIHT, University of Toulouse, and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research interests include statistical signal and image processing with a focus on Bayesian and Markov Chain Monte Carlo methods.

Dr. Tourneret has been involved in the EUSIPCO'02 (Program Chair), ICASSP'06 (Plenaries), SSP'12 (International Liaisons), CAMSAP'2013 (Local Arrangements), SSP'2014 (Special Sessions), and MLSP'2014 (Special Sessions). He was the General Chair of the CIMI Workshop on optimization and statistics in image processing hold in Toulouse in 2013 (with F. Malgouyres and D. Kouamé) and of CAMSAP 2015 (with P. Djuric). He has been a member of different technical committees, including the SPTM Committee of the IEEE Signal Processing Society, from 2001 to 2007 and from 2010 to 2015. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2008 to 2011. He has been an Associate Editor of the *EURASIP Journal on Signal Processing* since 2013.

Simon Godsill (M'95) joined CEDAR Audio Ltd., Cambridge, U.K., as a Founding Staff Member in 1988, where he is currently the Director. He is also the Professor of Statistical Signal Processing with the Department of Engineering, Cambridge University, Cambridge, U.K. He is also a Professorial Fellow and a Tutor with the Corpus Christi College Cambridge, Cambridge, U.K. He coordinates an active research group in signal inference and its applications within the Signal Processing and Communications Laboratory, Cambridge University, specializing in Bayesian computational methodology, multiple object tracking, audio and music processing, and financial time series modeling. He has authored journals, books, and international conference proceedings. He has co-authored a Springer text Digital Audio Restoration with Prof. P. Rayner in 1998.